

DOCUMENT RESUME

ED 196 943

TM 810 064

AUTHOR Harrington, Phillip J.
 TITLE Criticisms of Standardized Admissions Testing: A Historical Perspective.
 PUB DATE 24 Oct 80
 NOTE 14p.
 AVAILABLE FROM Paper presented at the Annual Meeting of the Northeastern Educational Research Association (Ellenville, NY, October 24, 1980).
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS College Entrance Examinations; Educational History; *Multiple Choice Tests; *Standardized Tests; Test Bias; *Testing Problems
 IDENTIFIERS *Criticism; *Testing Industry

ABSTRACT

An historical overview is presented of criticisms arising during an early period of large-scale standardized testing in the United States: the post World War II period when the Educational Testing Service was established. Not only is a brief description of the controversy provided, but the fact that many problems of assessment are not transient issues is demonstrated. These assessment problems continue to beset standardized testing and other attempts at measuring human behavior. Most criticisms of testing are rooted in basic value differences between the various publics and are not likely to be entirely resolved by education, legislation, or self-regulation by the industry. (Author/GK)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED196943

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Criticisms of Standardized Admissions
Testing: A Historical Perspective

Phillip J. Harrington
Cornell University

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

P. Harrington

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Paper presented at the Northeastern Educational
Research Association Annual Convocation
Ellenville, NY: October 24, 1980

TM 810 064

Criticism of Standardized Admissions
Testing: A Historical Perspective

The impetus for this paper originated when I was reviewing for another purpose Banesh Hoffmann's 1962 book The Tyranny of Testing. After considering his arguments and recommendations, it occurred to me that they were not very different than those of New York's "Truth in Testing" legislation or the Nader-Nairn report. Well, I was struck by the similarity, and being relatively naive to most of the historical arguments concerning testing, I decided to examine these earlier arguments a bit closer. These criticisms will sound very familiar to some, however realize that most of the present day criticisms have been around for well over 25 years. And, to appreciate the durability and saliency of the issues involved, all one needs to note is yesterday's "Truth in Testing" symposium here at NERA or tomorrow's ETS conference in New York City on "Admissions Testing and the Public Interest."

For purposes today I will pick up in the immediate post-WWII period when the previously autonomous College Board merged with the Carnegie Foundation and the American Council on Education to form the Educational Testing Service (ETS).

It was during this post-war period that college admissions testing and ETS grew at a tremendous rate. The returning soldiers applied to the colleges in record numbers and the development of high-speed computing made large scale testing a feasible and profitable enterprise.

By 1954, ETS was administering 850,000 tests per year with gross revenues topping \$3.5 million (Hale, 1955, p. 16). It was also during the mid-fifties that critics of standardized testing began voicing their sentiments concerning the ubiquitous nature of the tests. As one commentator noted, "whether your ambition is to be an ambassador or just a freshman at one of the 'better' colleges, the chances are that at some point you must first surmount the hurdles erected by ETS and its 'objective' tests" (Hale, 1955, p.14).

While people were beginning to criticize the tests on ground of their pervasive nature, many realized a dire need for efficient college admissions procedures. Some feared that by 1965 the college admissions process would be in a state of chaos similar to that experienced immediately following WWII unless an efficient admissions process was used (Bowles, 1957). Yet the critics continued to voice with increasing intensity their objections, resentments, and fears of the testing industry, and in particular ETS.

By 1960 the stage was set for what would be the first of many battles the testing industry would have with citizens representing a wide variety of backgrounds and interests.

The Critics and Their Criticisms

The most vocal of these critics was Banesh Hoffmann, a professor of mathematics at Queens College. He received degrees from Oxford and Princeton and then went on to work with Albert Einstein during the 1930's on fundamental studies of relativity at the Institute for Advanced Studies. He first became involved in the problems of testing when he criticized the selection methods used by the Westinghouse Science Talent Search in 1943. Later, the Ford Foundation arranged to have Hoffmann examine a complete ETS College Entrance Examination Board test. Hoffmann wrote ETS a 30 page critique of the test, and ETS subsequently asked him to become a staff consultant. He declined the offer in order that he might have a free hand in critiquing the field of educational testing (Hoffmann, 1961).

While being the most vocal critic of the time, Hoffmann was quick to point out that a multiple choice test from a reputable test publisher was "not mere sheets of haphazardly chosen questions hastily thrown together," but rather, a "burnished testing instrument of the most modern sort, painstakingly constructed and calibrated on scientific lines by a highly professional team of experts" (Hoffmann, 1962, p. 58).

However, Hoffmann and others questioned whether or not the test-makers through their "narrow professionalism, their faith in statistics, their

relative indifference to the powerful side effects of their activities, and the enormous impact of these activities on the lives of us all, are not already too powerful" (Hoffmann, 1962, p. 41). In short, Hoffmann felt that over confidence in objective tests could have "dangerous consequences, not only for education, but for the strength and vitality of the nation" (Hoffmann, 1961, p. 37).

As standardized testing became more widespread during the late 50's and early 60's, parents, college youth, employees, in short, almost everyone, became more concerned about how these new developments in testing would affect their lives and the lives of their children. Magazines such as Harper's, Commentary, and the Atlantic Monthly had all published articles dealing with the testing issue. Soon to follow were the more popular magazines such as McCall's, Life, and the New York Times Magazine. The testing controversy also made it to television on NBC's The Open Mind, which brought together Hoffmann, Henry Chauncey, then president of ETS, and Fred Hechinger, education editor for the New York Times. It is apparent, then, that the initial controversy over testing was in full swing by 1961.

Besides the perceived monopolistic qualities of ETS and the pervasiveness of their tests, criticisms of this era could be grouped into three main categories: (1) over-reliance on psychometric methods; (2) inability of multiple choice tests to assess higher level skills; and (3) bias against the superior student.

In terms of the perceived over-reliance on psychometric methods, one critic felt that, "our admissions tests...are pretty crude. Virtually all admissions testing today rests on a single intellectual breakthrough now nearly 70 year old" (Bowles, 1957, p.57). Another critic felt that many "advances" in psychometrics and statistics should be viewed with skepticism because of the tendency for different statistics to become faddishly popular and then be applied to "everything" indiscriminantly (Rothney, Danielson, and Heimann, 1959). For example, Hoffmann, in his usual scathing manner stated that, over-reliance on statistics "warps perspective." Further, if a person "worships statistics he will simplify, fractionalize,

distort, and cheapen in order to force things into a numerical mold" (1962, p.143). Additionally, Hoffmann warned that, statistics were not a "substitute for intellectual standards nor a shield against all types of criticism" (1959, p.201).

Yet the area that caused the most furor at the time was the perceived inability of multiple choice tests to assess higher level skills, which disadvantaged the superior student. This was Hoffmann's major criticism leveled on the test makers. As he notes in his book, The Tyranny of Testing:

How genuinely difficult, how worthy of first-rate minds, can questions be for which answers must be picked at the rate of one every minute or so...How deeply can such questions probe and still be machine gradable (p.89).

The tests favor the nimble-witted, quick-reading candidates who form fast superficial judgements (p.92).

The people rated highest according to statistically-based norms are apt to be the brilliant representatives of mediocrity at its pinnacle (p.141).

William Hale, (1955) writing in the Reporter noted that Henry Chauncey admitted that objective tests were not "ideally suited to people interested particularly in the arts" and that such a person's score may not be indicative of their true ability (p.17). Chauncey reiterated this point four years later in a letter to the American Scholar (1959) by stating, "present tests do not, of course, measure all of the qualities involved in scholarly work nor are they perfect measures of the qualities they do test" (p.533).

In the same letter, however, Chauncey broke his silence and took exception to Hoffmann's assertion that the tests penalize the brilliant students by questioning Hoffmann whether he had any "real" evidence to support his contention. And, in 1961, in response to Hoffmann's original 1961 Harper's article, Chauncey wrote that Hoffmann's method "is to ferret out questions he thinks would appear ambiguous to the exceptional student, and then assume that all potential geniuses will see them his way, score poorly, and be lost to society. He has never presented evidence to support his thesis" (Harper's, 1961, p.6).

And so the skirmishes over testing which started spreading in the early 1950's erupted into quite a battle by 1961. It was during this barrage of attacks on the testing industry that Hoffmann was preparing his coup de grace, to once and for all put the test makers in their place. He proposed in 1959 that he would put together a "critical analysis" of test items drawn from samples of College Board items and stated that:

...if to a significant extent, my critical analysis survives the rebuttal of the test experts, then, I will claim not only that there is prima facie evidence for the proposition that the College Board tests should be viewed with concern, but that if this concern is justified in the case of products of the Educational Testing Service, then one can tentatively assume that similar tests made by other organizations should also be viewed with concern....Moreover...I will claim that there is prima facie evidence for the need for a full scale inquiry into the whole field of testing for aptitude and ability" (1959, p.201-2).

It soon became clear that Hoffmann was not about to be intimidated by ETS. He felt that the way to penetrate the testing industry's defense was not to make general criticisms but to exhibit specific test questions, declare them defective, and challenge the test publishers to defend the specific questions. Further, he felt that challenging the content of specific items seemed "the one effective means by which the quality of multiple-choice tests can be called into question" (1961, p.42).

However, Hoffmann had a problem with his proposed approach. Since the content of the tests were not disclosed, he had no way to complete a "comprehensive and independent judgment" of the tests which would be made in the "public interest" (Hoffmann, 1961, p.42). Yet, he theorized, if there should be an "explicit and convincing public demonstration" that the test publisher's present methods produced dangerously defective questions, it would be difficult for them to "honorably and responsibly" ignore or hamper an independent public investigation (1961, p.42).

Hoffmann's master plan, as described in his American Scholar (1959) and Harper's (1961) articles and later in his book, The Tyranny of Testing (1962), was to establish a completely independent board of eminent educators

and scholars who would have access to the entire range of questions on the published tests. This board, drawn from organizations such as the American Council of Learned Societies, National Academy of Sciences, and the Modern Language Association, would have, as a minimum concern, the "quality" of multiple choice tests and their developers. For the benefit of the test users, Hoffmann proposed that this committee might draft a user "Bill of Rights" (Hoffmann, 1962, p.214).

Overall, Hoffmann felt that the committee's task should be to:

...formulate policy, develop recommendations for remedial action, and propose specific steps to implement its recommendations. In doing so it should consider the whole field of testing in all its aspects, with particular reference to the effects of tests on education, business, and the strength and vitality of the nation -- for these are the transcendent issues (1962, p.214-5).

While Hoffmann considered this plan the one effective means of controlling the testing industry's growth, others, while sympathetic to Hoffmann's criticisms, disagreed entirely with his method of attack.

One critic of Hoffmann was quite vivid when he stated:

Let us have a completely independent Board of eminent educators and scholars, preferably dentists, examine Dr. Hoffmann. If they should find as many as one or two cavities in any of Dr. Hoffmann's teeth, let's eliminate as rapidly as possible Dr. Hoffmann... (Harper's, 1961, p.8).

So during the period of 1959 - 62, when Hoffmann was espousing the merits of an external review of essentially all standardized testing, how was ETS reacting? President Chauncey of ETS agreed by stating in 1961, "that it might be well to have a high level commission that would look into the whole matter and make a fairly authoritative report which would give some feeling of confidence and assurance to the people who are having children undergoing tests, to the schools that are using tests, and to the pupils themselves (NBC, 1961, p.18). The following month he repeated his openness in Harper's by stating "we live by choice in a glass house and welcome 'investigation' by any competent individual or group" (1961, p.6).

Soon Hoffmann and followers had attracted a contingent of persons criticizing their own approach as being specious, and attacking the wrong cause. One critic (Wolfe, 1963) asserted that the tests were merely tools, and that the true responsibility lie with the user of such tests. Likewise another critic (Dyer, 1961) asserted that Hoffmann's approach was similar to saying that we should outlaw automobiles because of the potential for misuse by reckless drivers.

It soon became evident that Hoffmann was content to sit back and take pot-shots at the tests and test publishers. By doing this and failing to come up with any alternatives to the current practice of testing, he had left himself open to much criticism. As Gardner (1961) thoughtfully noted, "anyone attacking the usefulness of the tests must suggest workable alternatives" (p.49).

However, this is not to say the criticisms Hoffmann and others made of the testing industry in the 50's and early 60's had no positive effect. While one cannot be absolutely certain, many changes came about in the realm of testing during this period which may not have come about at all, or as rapidly, had people like Hoffmann been so vocal. For example, Rothney, et al. (1959) saw signs of progress in the area of testing during this period. They noted that some published tests indicated more careful attempts at standardization and validation; standards were raised for training and employment of guidance workers; test standards were published; and the provisions were made for a vehicle by which test reviews could reach the consumer.

Additionally, other groups began looking into their own activities. For example, during 1960 the North Central Association conducted an external review of its national testing activities. The study committee included members from the American Association of School Administrators, Council of Chief State School Officials, and the National Association of Secondary School Principals. The purpose of this committee testing operations was to examine the state of standardized testing in the United States and to present its findings to the Association. However, the committee was certain

that no progress would be made by attempting to issue ultimata or exercise control through force. They made it clear that they were not attempting to regulate or to control the testing industry (Vroman, 1961).

Indeed, most persons both inside and outside of the testing industry felt that legislation or external control would not be in the best interest of either producers or consumers of tests. Most of these same people felt that a combination of consumer education and self-regulation by the industry was the best route toward improving tests and their use (e.g., Rothney, et al., 1959).

The concern over the implications of standardized testing continued during the next few years and eventually culminated in the June, 1965 congressional hearings by both Senate and House committees. Most observers at the time felt this to be the most serious attack ever launched by citizen or government groups against any facet of psychological research or services. These hearings brought invited testimony from such persons as Arthur Brayfield, Karl Menninger, Sargent Shriver and Martin Gross (author of the Brain Watchers). The investigation and testimony involved not only testing for college admissions, but all aspects of psychological assessment.¹

By 1965, the need for increased self-regulation was apparent as Samuel Messick (1965), writing in the American Psychologist prophesied "we can anticipate that this public hostility to psychological testing may be funneled through institutional and legislative channels....In order to avoid these external controls it is argued by some that we should seize the initiative by moving toward self-regulation" (p.136).

However, while realizing the fallibility of the tests, Messick felt that many of the criticisms of testing were "blanket indictments" and what was needed was better public relations rather increased regulation. He went on to stress that the testing profession should not "refuse to implement further self-regulation, but rather that [they] should be

cautious and insist that any changes in ethical standards be based upon their intrinsic merits and not upon external pressures" (1965, p.137).

While the testing profession already had a set of self-regulatory standards, Technical Recommendations for Achievement Tests, published by AERA and NCME in 1955, two problems inhibited the documents use. First, there was no workable procedure for enforcing the accepted principles. And secondly, the standards lacked any clear hierarchy of precedence which reflected mainly a dissonance in terms of values among test-makers and the users of tests. Unfortunately, both factors reflect persistent problems in the case of self-regulation by a profession, and both problems continue to limit the use of the present Standards for Educational and Psychological Tests (1974).²

Messick concluded his article by recognizing that much of the testing controversy represents "a dialogue between conflicting values" (1965, p.141). Further, as mentioned above, "these value conflicts exist not only between the testers and the critics but within psychology as well" (1965, p.141). He felt that some of these value conflicts would have to be resolved by policy decision. Although, he warned, a policy decision must be made with extreme caution because any decision which would favor one side over the other, or which compromised an issue, "may create an atmosphere in psychology that would bias scientific and professional values of the future" (1965, p.141). Therefore, he felt if the "pressures of reality lead to establishment of further policy-based self-regulation, it would seem imperative to include at the same time formal provisions for its continuing reappraisal" (1965, p.141).

Summary and Conclusions

This paper presents an overview of criticisms arising during an early period of large-scale standardized testing in the United States. Its purpose was not only to provide a brief historical description of the controversy, but to demonstrate that many problems of assessment are not transient issues, but rather continuing problems which beset standardized testing and perhaps all attempts at measuring human behavior.

For example, all one has to do is compare the obvious similarities of Hoffmann's concern that standardized tests are biased against the superior student, with later and continuing criticisms of test bias toward special groups. Or, Hoffmann's call for an examination of specific test items with enacted or proposed truth-in-testing legislation. Or, early critics arguments on the lack of predictive validity, with many arguments of the Nader/Nairn (Nairn, 1980) report. The comparisons could go on and on. The point is, I feel that most of these criticisms are rooted in basic value differences between the various publics and are not likely to be resolved entirely by any amount of legislation, education, or self-regulation by the industry.

Testing in America is once again under attack. This time, however, the critics have mustered the support of the nation's lawmaking bodies. Advances in psychometrics have been largely ineffectual in convincing the public and lawmakers of the propriety and benefits of testing, as has increased and improved public relations by the testing industry. As Gardner writes in his book Excellence (1961), "the tests are designed to do an unpopular job. Even if these sources of concern were to disappear, the hostility toward the tests would probably remain. As the tests improve and become less vulnerable to present criticism, the hostility to them may actually increase" (p.47).

Reference Notes

1. For an excellent account of the 1965 congressional investigation into testing, the reader is referred to a special issue of the American Psychologist 20 (11) November, 1965.
2. For an excellent discussion of the standard setting and enforcement process within a professional field see Ridings, J.M. Standard Setting in Accounting and Auditing: Considerations for Educational Evaluation. Unpublished Ed.d. dissertation, Kalamazoo: Western Michigan University.

Bibliography

- American Educational Research Association, National Council on Measurement in Education. Technical Recommendations for Achievement Tests. Washington, D.C.: National Education Association, 1955.
- American Psychological Association, American Educational Research Association, National Council on Measurement in Education. Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, 1974.
- American Scholar, 28, No. 4, Autumn, 1959.
- Bowles, Frank H. Candidates and confusion. College and University, 1957, 32, 467 - 480.
- Dyer, H.S. Is testing a menace to education? New York State Education, 1961, 16 - 19.
- Gardner, John W. Excellence. New York: Harper and Row, 1961.
- Hale, W.H. The quizmasters fasten onto higher education. The Reporter, 1955, 13, 14 - 20.
- Harper's, May 1961. 222.
- Hoffmann, B. "Best answers" or "better minds?" American Scholar, Spring 1959, 28, 195 - 202.
- Hoffmann, B. The tyranny of multiple choice tests, Harper's, March 1961, 222, 37 - 44.
- Hoffmann, B. The Tyranny of Testing. Crowell-Collier Press, 1962.
- Messick, S. Personality measurement and the ethics of assessment. American Psychologist, 1965, 20, No. 2.
- Nairn, A. and Associates. The Reign of ETS: The Corporation That Makes Up Minds. Washington, D.C., 1980.
- National Broadcasting Corp. The Open Mind. April 2, 1969 (Transcript).
- Rothney, J.W.M., Danielson, P.J., Heimann, R.A. Measurement for Guidance. New York: Harper & Brothers, 1959.
- Vroman, C. First report on testing by external agencies. North Central Association Quarterly, 1961, 35, No. 2, 223 - 225.
- Wolfe, D. Education tests. Science, December 20, 1963, 142, No. 3599, 1.